ST304 Week 9

Kaixin Liu[1]

[1]PhD Student in Statistics, LSE

Mar 19, 2026

Table of Contents

Table of Contents

Least squares estimation

Let $\omega = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^\top$ and $r = \max(p, q)$.

Conditional least squares

Define the recursive residuals by

$$e_t(\omega) = X_t - \sum_{j=1}^{p} \phi_j X_{t-j} - \sum_{k=1}^{q} \theta_k e_{t-k}(\omega), \qquad e_t(\omega) = 0 \text{ for } t \leq 0.$$

Then

$$S_n(\omega) := \sum_{t=r+1}^{n} e_t(\omega)^2, \qquad \widehat{\omega}_{LS} \in \arg\min_{\omega} S_n(\omega).$$

Remarks

- For AR(p), $S_n(\omega)$ is quadratic and $\widehat{\omega}_{LS} = (Z^\top Z)^{-1} Z^\top Y$.
- For general ARMA, $S_n(\omega)$ is not quadratic in the MA parameters, so numerical optimisation is needed.
- Under regularity conditions, least squares is consistent and asymptotically equivalent to MLE.

Maximum likelihood estimation

Let $r = \max(p, q)$ and assume $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

Gaussian likelihood

Write

$$\widehat{X}_{t|t-1} := E(X_t \mid X_{t-1}, \dots, X_1), \qquad \sigma^2_{t|t-1} := \mathrm{Var}(X_t \mid X_{t-1}, \dots, X_1),$$

and $e_t := X_t - \widehat{X}_{t|t-1}$. Then

$$L_n(\omega, \sigma^2) = f_{\omega,\sigma^2}(X_1, \dots, X_r) \prod_{t=r+1}^{n} (2\pi\sigma^2_{t|t-1})^{-1/2} \exp\left(-\frac{e_t^2}{2\sigma^2_{t|t-1}}\right).$$

Remarks

- Conditional MLE drops the initial density term $f_{\omega,\sigma^2}(X_1, \dots, X_r)$.
- For AR(p), conditional MLE coincides with least squares.
- In practice, MLE is obtained numerically and usually constrained to the causal / invertible region.

Table of Contents

Likelihood ratio tests, AIC, BIC and cross-validation

Model comparison

Likelihood ratio tests compare **nested** models. AIC and BIC balance fit and complexity:

$$\mathrm{AIC} = -2\log L_n(\widehat{\omega}, \widehat{\sigma}^2) + 2\dim\Theta,$$
$$\mathrm{BIC} = -2\log L_n(\widehat{\omega}, \widehat{\sigma}^2) + \log(n)\dim\Theta.$$

Cross-validation chooses the model with the smallest out-of-sample validation error.

How to count $\dim\Theta$

For mean-zero ARMA(p, q) with free $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2$,

$$\dim\Theta = p + q + 1.$$

If μ is also unknown, then $\dim\Theta = p + q + 2$. Equality constraints reduce the dimension; e.g. AR(2) with $\phi_2 = \phi_1^2$ has dimension 2 if $\mu = 0$.

Remark. Causality / invertibility are inequality restrictions, so they do not usually reduce $\dim\Theta$.

Table of Contents

Ljung–Box test

Portmanteau test

Using residuals $\widehat{\varepsilon}_t$ from a fitted ARMA(p, q) model, the Ljung–Box statistic is

$$Q_{n,m} = n(n+2) \sum_{k=1}^{m} \frac{\widehat{\rho}_{\widehat{\varepsilon}}(k)^2}{n-k},$$

where $\widehat{\rho}_{\widehat{\varepsilon}}(k)$ is the residual sample ACF. It tests

$$H_0 : \rho_{\widehat{\varepsilon}}(1) = \cdots = \rho_{\widehat{\varepsilon}}(m) = 0.$$

Interpretation

If the ARMA(p, q) fit is adequate and $m > p + q$, then

$$Q_{n,m} \approx \chi^2_{m-p-q}.$$

A large value (small p-value) indicates residual autocorrelation remains.

Dickey–Fuller test

AR(1) unit-root test

For $X_t = \phi X_{t-1} + \varepsilon_t$, write

$$\Delta X_t = (\phi - 1)X_{t-1} + \varepsilon_t = \delta X_{t-1} + \varepsilon_t.$$

Then test

$$H_0 : \delta = 0 \ (\phi = 1,\ \text{unit root}) \qquad \text{vs} \qquad H_1 : \delta < 0 \ (|\phi| < 1).$$

Remarks

- Under H_0, the test statistic has a **non-standard** Dickey–Fuller limiting distribution, so usual t critical values are invalid.
- The augmented Dickey–Fuller test generalises this to AR(p) by using

$$\Delta X_t = \delta X_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta X_{t-j} + \varepsilon_t.$$

Table of Contents

MSPE and best h-step-ahead forecast

Forecast target

For a predictor $f(X_n, \dots, X_1)$, define

$$\mathrm{MSPE}(f) := E\left[\left(X_{n+h} - f(X_n, \dots, X_1)\right)^2\right].$$

The best linear h-step-ahead forecast is

$$\widehat{X}_{n+h|n} := \Pi(X_{n+h} \mid X_n, \dots, X_1) = \arg\min_{f \text{ linear}} \mathrm{MSPE}(f),$$

Conditional expectation

If the model is causal with independent innovations, then

$$E(X_{n+h} \mid X_n, \dots, X_1)$$

minimises MSPE over *all* measurable predictors. Under Gaussianity, it equals the best linear predictor.

ℓ-step prediction variance

General form

If

$$X_{n+\ell} = \sum_{j=0}^{\infty} \lambda_j X_{n-j} + \sum_{j=0}^{\ell-1} \psi_j \varepsilon_{n+\ell-j},$$

then

$$\widehat{X}_{n+\ell|n} = \sum_{j=0}^{\infty} \lambda_j X_{n-j}, \qquad \mathrm{Var}(X_{n+\ell} - \widehat{X}_{n+\ell|n}) = \sigma^2 \sum_{j=0}^{\ell-1} \psi_j^2.$$

Example: AR(1)

For $X_t = \phi_1 X_{t-1} + \varepsilon_t$ with $|\phi_1| < 1$,

$$\widehat{X}_{n+\ell|n} = \phi_1^{\ell} X_n, \qquad \mathrm{Var}(X_{n+\ell} - \widehat{X}_{n+\ell|n}) = \sigma^2 \frac{1 - \phi_1^{2\ell}}{1 - \phi_1^2}.$$

As $\ell \to \infty$, the prediction variance tends to $\mathrm{Var}(X_t)$.

Recursive forecasting

Step 1: estimate past errors

Start with $\widetilde{\varepsilon}_t = 0$ for $t \leq 0$. For $t = 1, \ldots, n$, compute

$$\widetilde{\varepsilon}_t = X_t - \sum_{j=1}^{p} \phi_j X_{t-j} - \sum_{k=1}^{q} \theta_k \widetilde{\varepsilon}_{t-k}.$$

Step 2: predict future values

For $t > n$, set $\widetilde{\varepsilon}_t = 0$ and use

$$\widetilde{X}_t = \sum_{j=1}^{p} \phi_j \widetilde{X}_{t-j} + \sum_{k=1}^{q} \theta_k \widetilde{\varepsilon}_{t-k}, \qquad \widetilde{X}_s = X_s \text{ for } s \leq n.$$

In practice, replace ϕ_j, θ_k by their estimates. For AR(1), this reduces to

$$\widehat{X}_{n+\ell|n} = \phi_1 \widehat{X}_{n+\ell-1|n}.$$